

CLAIMS

I claim:

1. A method of managing compute resources within a compute environment, the method comprising:

identifying a need type and a group of available compute resources;

creating a reservation mask over the identified group of compute resources; and

if a request from a consumer matches the need type, then constraining the creation of a sub-reservation for the consumer to only use compute resources within the reservation mask.

2. The method of claim 1, further comprising:

creating a set of reservation masks over the identified group of compute resources, wherein multiple sub-reservations created from multiple consumer requests are each constrained to only use compute resources within the set of reservation masks.

3. The method of claim 1, wherein the request from the consumer is a non-administrator request.

4. The method of claim 3, wherein an administrator is not constrained within the reservation mask.

5. The method of claim 1, wherein the compute environment is a grid environment and the request from a consumer is a grid-based request.

6. The method of claim 1, wherein the compute environment is one of: a compute farm, a data center, a virtual hosting center, a hosting center, a grid and a cluster.

7. The method of claim 1, wherein the request from a consumer is a personal-based request.

8. The method of claim 1, wherein the request from a consumer is placed in an access control list.

9. The method of claim 1, wherein if a request from the consumer does not match the need type, then no constraints are enforced for creating the sub-reservation.
10. The method of claim 1, wherein creating the reservation mask further comprises specifying at least one timeframe during which the reservation mask enforces constraints.
11. The method of claim 10, wherein the at least one time frame further comprises a plurality of independent time frames.
12. The method of claim 10, wherein the at least one time frame further comprises a plurality of regular, periodic timeframes.
13. The method of claim 1, further comprising:
specifying an access control list that constrains which consumers or resource requests may utilize compute resources within the reservation mask.
14. The method of claim 1, wherein the need type comprises at least one of: a particular use, a user, a group of users, a job source and a type of job submission.
15. The method of claim 1, wherein the reservation mask is for a need type of a personal reservation or a grid reservation.
16. The method of claim 15, wherein the personal reservation comprises a reservation that dedicates resource access to a user or a group of users.
17. The method of claim 16, wherein if the personal reservation provides access to resources to a group of users, then each reservation and reservation timeframe are determined by a user in the group of users that requests the respective reservation.
18. The method of claim 15, wherein a grid reservation is a reservation requested from outside an administrative group.

19. The method of claim 1, wherein the step of identifying a need type and a group of available resources is based on an administrative policy.
20. The method of claim 1, further comprising:
modifying the sub-reservation according to received data.
21. The method of claim 20, wherein the received data is at least one of resource usage, system performance, a policy and a criterion associated with the request.
22. The method of claim 20, wherein modification of the sub-reservation is bounded by a minimum threshold and a maximum threshold.
23. The method of claim 20, wherein modifying the sub-reservation further comprises modifying at least one of: an access control list, reserved resources and a time frame covered.
24. The method of claim 1, wherein the created sub-reservation is constrained by independent, non-administrative criteria.
25. The method of claim 24, wherein the criteria comprises at least one of quantity of resources and on a per-credential basis.
26. The method of claim 1, wherein the reservation mask is a policy enforcing mechanism to manage and constrain sub-reservations.
27. A method of managing compute resources within a compute environment, the method comprising:
identifying a need type and a group of available resources;
creating a reservation mask over the identified group of resources; and

if a request from a requestor matches the need type, then constraining the creation of a sub-reservation for the requestor independent of the reservation mask and according to credentials associated with the request.

28. The method of claim 27, wherein the credentials are one of: a per user credentials, per group credentials, per class credentials, quality of service-based credentials and a partition-based credentials.

29. The method of claim 27, wherein the request is a non-administrative request.

30. The method of claim 29, wherein the request is a request for a sub-reservation comprising one of a grid-based reservation and a personal reservation.

31. A system for managing compute resources within a compute environment, the system comprising:

means for identifying a need type and a group of available resources;
means for creating a reservation mask over the identified group of resources; and
means for constraining the creation of a sub-reservation associated with a request from a consumer that matches the need type.

32. The system of claim 31, wherein the means for constraining the creation of a sub-reservation constrains the creation of the sub-reservation to only use compute resources within the reservation mask.

33. The system of claim 31, wherein the means for constraining the creation of a sub-reservation constrains the creation of the sub-reservation independent of the reservation mask and according to credentials associated with the request.

34. The system of claim 33, wherein the credentials are one of: a per user credentials, per group credentials, per class credentials, quality of service-based credentials and a partition-based credentials.

35. A system for managing compute resources within a compute environment, the system comprising:

- a module configured to identify a need type and a group of available resources;
- a module configured to create a reservation mask over the identified group of resources; and
- a module configured to constrain the creation of a sub-reservation for the consumer if a request from a consumer matches the need type.

36. The system of claim 35, wherein the module configured to constrain the creation of the sub-reservation constrains the creation of the sub-reservation to only use resources within the reservation mask.

37. The system of claim 35, wherein the module configured to constrain the creation of the sub-reservation constrains the creation of the sub-reservation independent of the reservation mask and according to credentials associated with the request.

38. The system of claim 37, wherein the credentials are one of: a per user credentials, per group credentials, per class credentials, quality of service-based credentials and a partition-based credentials.

39. A computer-readable medium storing instructions for controlling a computing device to manage compute resources within a compute environment, the instructions comprising:
identifying a need type and a group of available resources;
creating a reservation mask over the identified group of resources; and
constraining the creation of a sub-reservation for the consumer if a request from a consumer matches the need type.

40. The computer-readable medium of claim 39, wherein the sub-reservation is constrained to only use resources within the reservation.

41. The computer-readable medium of claim 39, wherein the sub-reservation is created for the request independent of the reservation mask and according to credentials associated with the request.